



Citation for published version:

Day, M 2003, Digital preservation and long-term access to the content of electronic serials. in W Jones (ed.), *E-series: publishers, libraries, users, and standards, 2nd ed.*. Haworth Information Press, Binghamton, NY, pp. 167-195.

Publication date:
2003

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Digital preservation and long-term access to the content of electronic serials

Michael Day

Abstract: This paper is an investigation of digital preservation issues that are related to the development and use of scholarly online electronic serials (e-serials). Firstly, some background topics are discussed. These include: a discussion of what preservation means in the context of e-serials; a look at the current and future development of scholarly communication; and an outline of the major formats currently used for the distribution of e-serials. Digital preservation issues are then discussed in more detail, including an examination of the technical difficulties of preserving information in digital form. The paper then considers some preservation strategies that have been proposed, including technology emulation and data migration and the role of metadata. There follows a brief introduction to some important non-technical issues. These include: the difficulty of ensuring the authenticity of digital resources; the changing intellectual property rights context; and a discussion of who should have responsibility for preserving e-serials and the criteria that could be used for selecting digital content for preservation.

Keywords: e-serials; online serials; electronic journals; digital libraries; networks; digital preservation; digital archives

Michael Day is Research Officer at the UK Office for Library and Information Networking (UKOLN), University of Bath, Bath BA2 7AY, United Kingdom (email: m.day@ukoln.ac.uk).

UKOLN is funded by the Joint Information Systems Committee of the UK higher and further education funding councils and by Resource: the Council for Museums, Archives and Libraries, as well as by project funding from several sources. The views expressed in this paper do not necessarily reflect those of UKOLN or its funding bodies.

Introduction

The long-term preservation of information in digital form has been a subject of growing interest for the library and information professions since the late 1970s.

Initially, this was a response to the growing use of digital technologies in publishing and the suggestion by Lancaster (and others) that society was steadily moving towards the use of information systems that would be largely paperless.¹ Much recent awareness of the digital preservation issue was facilitated by the publication in 1996 of the report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access (CPA) and the Research Libraries Group (RLG).² This report provided a good summary of relevant issues and acted as a catalyst for further research and development.

Naturally, digital preservation issues apply across the whole range of digitally based information resources that exist. This paper will be concerned with just one subset of these; i.e. the digital products of scholarly and scientific communication that have been traditionally published in printed peer-reviewed serials. In order to reflect current trends, this will include both the contents of those electronic serials that have been formally published by commercial and learned society publishers and also those papers (or e-prints) that have been "self-archived." The development of sustainable preservation strategies for both of these types of resource is essential to ensure the future viability of scholarly communication.

Initially, however, it is important to define some terms. Even the concept of "digital preservation" can be difficult to define unambiguously. For example, the term

is sometimes confused with the quite separate idea of digitising resources as part of a preservation strategy for non-digital objects. In the context of scholarly communication, back issues of serials are often digitised and made available over networks in order to improve access to the information contained in them and, on occasion, to aid the preservation of the original item.³ Examples of this type of service are JSTOR and the Internet Library of Early Journals (ILEJ).⁴ Although it is acknowledged that digitised versions of the back issues of serials will also need to be preserved, they will not form the main focus of this paper.

The concept of digital preservation can also get muddled with ideas of digital archiving or digital archives. Indeed, the terms sometimes appear to be used interchangeably. This can be a source of confusion. For example, members of the archives and records professions will have a very different understanding of the archives concept than computer scientists, who often use the word "archive" as a verb, meaning the creation of secure backup copies for a fixed period of time. The word is used in a similar way by the developers of e-print archives. Wherever possible, this paper will use the word "preservation" in an attempt to avoid any confusion, except in relation to e-print archives. Preservation itself is primarily concerned with the survival of information in a usable form for as long as it is required.⁵ Preservation, therefore, is not just concerned with the conservation or restoration of physical artefacts, but includes all of the strategic and organisational considerations that relates to the survival of information over time. Hedstrom has usefully defined digital preservation as "the planning, resource allocation, and application of preservation methods and

technologies to ensure that digital information of continuing value remains accessible and usable."⁶

A distinction is sometimes made between the preservation of the information embodied in a document (the information content) and the physical conservation of an information carrier.⁷ This is especially relevant for digital information - including the content of e-serials - because most users tend not to be interested in which particular type of physical object is being used for storing the information. There may be a case for retaining physical objects, for example in a museum of technology, but for e-serials, the preservation of informational content is the main issue.

Another important, often emotive, issue is how long this information should be kept. It is often assumed that preservation should be permanent, often defined with reference to loaded terms like "in perpetuity" and "indefinitely." In the context of archives, Bearman has noted the absurdity of using concepts of permanence with regard to preservation and instead has proposed a more realistic concept of "retention for period of continuing value."⁸ In consequence, this paper will assume that preservation is normally linked to continuing access and use and not with nebulous concepts of permanence.

The electronic serial landscape

Scholarly and scientific papers published in peer-reviewed printed serials have fulfilled a vitally important role in the scholarly communication process since the seventeenth century. Line has correctly described them as the "established medium of

record and dissemination."⁹ Despite this, there has been a growing feeling that the traditional printed serial has outlived its historical role and that some kind of digitally based form of scholarly communication will soon replace it.¹⁰ At the moment, there are two main models for the digital distribution of scholarly communication through the Internet, and both have developed relatively independently.¹¹ Firstly, the publishers of traditional printed serials have begun to place digital copies of each issue on Web sites (a form of parallel publishing) and have started to experiment with the production of digital-only journals. Secondly, some scholars and librarians have outlined "subversive proposals" that aim to ensure free access to the scholarly and scientific literature through the self-archiving of papers by their authors. Both the publishers of e-serials and the supporters of author self-archiving initiatives hope - in different ways - that developments in the digital distribution of scholarly communication might help offset the effects of the ongoing "serials crisis."¹²

The functions of serials

The idea of electronic serials has been around for a number of years. Research projects in the 1980s first proved that e-serials were technically feasible.¹³ Initiatives like EIES (Electronic Information Exchange System) and BLEND (Birmingham and Loughborough Electronic Network) demonstrated that all stages of the production of a scholarly serial, (i.e. article submission, peer-review, editing and distribution), could be undertaken through computer networks.¹⁴ However, these projects did not develop into sustainable services at the time because computer networks were immature and not ubiquitously available, even within the research community.¹⁵ Also, in some cases,

the proposed user-interfaces were poor and the system developers did not always have a clear idea of the many different functions fulfilled by the printed scholarly serial. McKnight has noted that there is no incentive for either authors or readers to change unless electronic serials can do "at least the same things - and preferably more - with electronic journals as they do with paper."¹⁶ The printed peer-reviewed serial has proved to be an extremely successful part of scholarly communication since the seventeenth century because it has fulfilled a wide range of different functions. The following list is expanded from those functions that have been identified by Rowland.¹⁷

- *Dissemination* - publication of a paper in a peer-reviewed serial allows an author to disseminate important research findings to the wider research community and beyond. It is important to recognise, however, that the content of many published papers may also have been discussed informally, reported on at conferences or distributed as pre-prints.
- *Quality control* - consistently applied editorial processes can help to ensure a high written standard of papers, but the main quality control process is the peer-review of all submitted papers.
- *Establishing priority* - one of the most important functions of the printed serial, especially in the science, technology and medicine (STM) disciplines, is to be able to establish priority over a particular discovery or advance.

- *The recognition of authors* - authors value publication in refereed-serials as a means of raising their profile and as a means of gaining further research contracts or promotion.
- *The creation of a public domain "archive"* - once published, serial papers are in the public domain and research libraries collectively act as a distributed "archive," preserving the knowledge embodied in them for future scholars.

These are the basic functions that will need to be fulfilled by any new form of digitally based scholarly communication.

Electronic serials

Since the late 1980s, both commercial and not-for-profit publishers (e.g. learned societies) have invested significant amounts of time and effort into the development of electronic serials. Early efforts included the CD-ROM-based experiment in document delivery undertaken by the ADONIS consortium. This was followed by initiatives like the American Chemical Society's Chemical Journals Online (CJO) service and collaborative ventures with libraries like the ELVYN project that involved the Institute of Physics Publishing and a research team based at Loughborough University.¹⁸ Another example of library-publisher co-operation was TULIP (The University Licensing Program), in which selected serials from the Elsevier Science group were delivered to participating US research libraries in an

attempt to investigate some of the technical, legal and economic issues associated with e-serials and user behaviour.¹⁹

It was, however, the emergence of the Internet as a mass medium that finally persuaded both commercial and learned society publishers that the development of electronic serials was a viable option. Publishers have started to distribute the content of their scholarly serials through the Internet. At the present time, probably for economic reasons, most publishers are using digital technologies in an essentially conservative way.²⁰ Most are creating Web-based services that give access to the content of already published printed serials rather than introducing new titles that implement digital-specific features like embedded multimedia. Examples of these services include ScienceDirect from Elsevier Science, the IDEAL Online Library from Harcourt (including Academic Press) and Project MUSE from the Johns Hopkins University Press. Many publishers "bundle" collections of e-serials together and offer institutions (and their libraries) the ability to subscribe to their entire list as a single product.²¹ Halliday and Oppenheim note that this helps publishers to spread production costs across the whole subscriber base and can help reduce the scope of license negotiations.²² At the same time, the organisations that subscribe to serials are also beginning to join together in order to collectively negotiate licenses with publishers. Examples include OhioLINK (a consortium of Ohio libraries) and the UK National Electronic Site Licence Initiative (NESLI). NESLI, a consortium led by the University of Manchester and Swets Blackwell, negotiates deals with publishers on

behalf of the whole UK higher education community, and where possible, incorporates a clause for ensuring long-term access.²³

Some research libraries are beginning to support partnerships with learned society publishers in order to provide digital outlets for existing titles and help foster the publication of new electronic serials. For example, HighWire Press - a not-for-profit initiative of the Stanford University Libraries - gives access to a large number of serials published by learned societies. These include *Science*, *Proceedings of the National Academy of Sciences* and *BMJ*. Another library-led initiative is SPARC (Scholarly Publishing and Academic Resources Coalition), a coalition supported by the Association of Research Libraries (ARL). This organisation looks for partnerships with learned society publishers and seeks to underwrite new serials that might be able to provide competition for more established and expensive titles. The first new titles published by SPARC partners included *Organic Letters* (published by the American Chemical Society), *PhysChemComm* (published by the Royal Society of Chemistry) and *Evolutionary Ecology Research*.²⁴

It is important to realise that most commercial and learned society publishers have not yet moved to digital-only publication. Many tend to offer access to parallel digital versions of printed articles - usually in PDF (Portable Document Format) - and some bundle both printed and digital formats into the subscription price. On occasion, however, the digital version will offer more functionality than the printed version, e.g. Springer's LINK service includes some papers that contain supplementary material like software, multimedia or data sets that would not be available in either of the

printed or PDF versions. The Web sites of some serials also become a focus of more informal communication and sometimes provide space for news items, scholarly debates, information about relevant events, links to external Internet resources, etc. The Web sites of general scientific serials like *Nature* and *Science* are good examples of this. It is important to recognise that the Web sites of serials themselves may also be good candidates for long-term preservation.

Self-archiving initiatives

At the same time as this mainly publisher-led activity has been unfolding, other actors in the scholarly communication chain have seen the potential for more wide ranging changes in scholarly communication, often suggesting that printed serials have no long-term future.²⁵ Some scholars and librarians have begun to ask why the *status quo* in paper serials should simply duplicate itself in the Internet age.²⁶ People have begun to support "subversive proposals" that suggest that the authors of scholarly and scientific papers should simply make them available for free by storing (or archiving) copies on Internet sites.²⁷

Proponents of the self-archiving idea argue from the premise that peer-reviewed serials form a peculiar type of publishing that has little in common with the payment or royalty based trade-publishing sector. So, for example, Harnad and Hemus argue that the authors of scholarly and scientific papers are not primarily interested in monetary reward, but in having their work read, used, built-upon and cited.²⁸ In the print world, the authors of papers had to perpetuate what Harnad has

called a "Faustian bargain," whereby they traded the copyright of works to publishers in exchange for having them printed and distributed. He argues that this type of bargain made sense when publishing remained an exclusive and expensive domain, but that it has little or no relevance in the Internet age when scholars and scientists can self-archive their own papers at little or no personal cost.²⁹ Authors, then, are being encouraged to deposit digital copies of their papers in centralised e-print services or in e-print archives based at their own institutions, thus making them freely available to fellow scholars and scientists. Furthermore technologies have now been developed that would enable distributed e-print archives implemented in accordance with the standards developed as part of the OAI (Open Archives Initiative) to be combined into a single global virtual archive.³⁰ It is argued that once users prefer to access the free online copy of papers, publishers' revenues will decline and their role eventually will be reduced to the essential one of providing quality control: chiefly the implementation of peer-review. This could then be paid for directly by authors' institutions out of savings on subscription costs.

The most frequently cited model of the first stages of the "subversive-proposal" in action is the e-print archive first set up by Ginsparg at the Los Alamos National Laboratory in 1991.³¹ This service originally gave access to digital pre-prints in the domain of high-energy physics and it very quickly became the primary means of scholarly communication in this subject area. It has since expanded to cover other areas of physics, mathematics and computer science. A physicist was quoted in 1994 as saying that the service had completely changed the way people in his field

exchanged information. He noted that the only time that he needed to look at published serials was to find articles that predated the Los Alamos physics databases.³² The success of the Los Alamos e-print archive has led to the development of other Web-based e-print services. This has not been without controversy. For example, in 1999, the US National Institutes of Health published a proposal for a service called PubMed Central that would give free online access to published material in the biomedical sciences.³³ The original proposal suggested the creation of two separate services: one that would publish papers with peer-review from the editorial board of participating serials and a second one for non-peer-reviewed papers - essentially a form of e-print archive. The non-peer-reviewed section of the proposed service received considerable criticism, largely focussed on the need for the strict evaluation of clinical research, because of potential adverse impacts on public health and medical practice.³⁴ Consequently, the PubMed Central system, when it was officially launched in January 2000, only contained the peer-reviewed part of the proposed service. Serials currently participating in PubMed Central include the *Proceedings of the National Academy of Sciences*, *Molecular Biology of the Cell* and the *BMJ*.

To date, many scholars and scientists have been reluctant to commit themselves to self-archiving. Part of this is due to cultural differences between different subject areas. Valauskas, for example, notes how different the styles of communication and verification, debate and consensus can be amongst different academic disciplines.³⁵ It is rather simplistic to say that because self-archiving appears

to be accepted by most physicists then it should also be adopted by biomedical researchers or ancient historians. Another possible reason why self-archiving has failed to take off is that scholars and scientists have been reluctant to stop publishing in established high-impact serials. Some supporters of the author self-archiving approach argue that there remain "perverse incentives" for scholars to publish their papers in expensive serials.³⁶ For example, the market for scholarly serials can be skewed by the fact that the organisations that actually spend money on subscriptions tend not to be the ones who actually read or submit articles to serials. For these reasons, Harnad and his supporters do not suggest that authors should completely give up publishing in high impact serials, but only that they simultaneously self-archive copies of the same papers either in a centralised e-print service or on their institution's own e-print archive. They argue that this in itself will be enough to trigger the inevitable change.

Following the launch of the PubMed Central service, some scholars and scientists have stepped up their campaign for the creation of e-print archives of published papers. A number of scientists have appealed to serial publishers in the life sciences to co-operate with initiatives like PubMed Central by making their content available to publicly accessible e-print archives some time after publication.³⁷ In order to help bring this about, a group known as the "Public Library of Science" has recently been inviting scholars to sign an open letter. Signatories pledge their intention only to publish in, undertake peer-review for and serve on the editorial boards of those serials that will make papers freely available six months after

publication. This initiative moves far beyond a general support for author self-archiving initiatives. It is using authors to put pressure on the publishers of serials to "give away" content to publicly-funded e-print initiatives. The Public Library of Science group argue that this will help to facilitate free access to the scientific literature but also suggest that open e-print archives can continue the historical role of research libraries with regard to preservation. The open letter states that "the permanent, archival record of scientific research and ideas should neither be owned nor controlled by publishers, but should belong to the public, and should be freely available through an international online public library."³⁸ This is a laudable aim, but the letter does not address precisely how this digital archival record of research would be preserved and how the preservation process would be co-ordinated. This is a weakness with most self-archiving models, which sometimes appear to treat long-term digital preservation as a relatively simple technical problem.³⁹ This is probably because the supporters of self-archiving publishing models tend to be concerned more with the distribution of current content than with ensuring continuing access to the scholarly record.⁴⁰

In any case, it is probably true to say that the digital distribution of scholarly and scientific papers through either e-serials or e-print archives will not fulfil all of the functions that have been fulfilled by printed serials. So, for example, while the digital distribution of papers is extremely good for fast and timely dissemination, the volatility of digital information means that it has not so good at fulfilling the functions of establishing priority and the long-term preservation of the scholarly record.

Meadows has noted that establishing priority is often the "basic motivation" of scientists, and that many consider it to be much more important than being read or cited by their peers.⁴¹ Establishing priority in a digital environment will depend upon the implementation of secure time-stamping and other authenticity mechanisms. This is just one aspect of a wider digital preservation problem that will need to be addressed by the publishers of e-serials, the organisers of centralised e-print services like PubMed Central and by institution-based interoperable e-print archives.

Distribution methods and formats in use

Electronic serials have been made available in a variety of different formats. The first e-serials used relatively simple formats like plain ASCII text or bit-mapped page images. ASCII was considered to be fine for articles that were comprised largely of text, but was not thought adequate for the representation of the visual complexity of some scholarly literature. With this in mind, the early publisher-led ADONIS project, for example, scanned the paper copy of the serial and distributed the pages as bit-mapped images on CD-ROMs.⁴²

The most difficult, and therefore expensive, types of serial to produce in any format are STM serials because they contain specialised terminology and frequently include detailed mathematical formulae, complex artwork or tabular data.⁴³ For these reasons networked STM serials often tend to use formats that retain the features of print serials. The most popular of these formats are PostScript and the proprietary Portable Document Format (PDF).⁴⁴ PDF is particularly good for use in situations

where digital versions of printed serials are being made available online, as the pages will look the same as in the printed version. Many commercial publishers distribute parallel editions of printed serials in PDF, largely because the format is very easy to produce as a by-product of the publication process for print, but also because they can maintain complete control over the page layout of their digital publications.⁴⁵

Widespread use of the Web as a distribution technology has meant that the other popular format currently used for the distribution of e-serials is the HyperText Markup Language (HTML). E-serials often use HTML because they can take advantage of the hypertextual and multimedia features of Web publishing rather than just replicating the printed page. For example, the UK serial *Internet Archaeology* has published a paper on Roman amphorae that contains "clickable" maps and timelines.⁴⁶ Many e-serials that were originally distributed in ASCII form by electronic mailing list software are now also available in HTML on the Web. HTML is not always an ideal format for STM serials as it has limitations in encoding some special characters and relies on inline graphics or helper applications for the full display of illustrations.⁴⁷ Accordingly, HTML is often used to create an interface for the viewing of other formats. Many commercial e-serial providers give contents page information and abstracts in HTML while the full-text of the articles are made available in PDF.

Naturally, there is an interest in other formats - especially those based on logical content mark-up, e.g. those based on the Standard Generalised Markup Language (SGML) and the Extensible Markup Language (XML). Project ELVYN, for example, had first considered PostScript as a delivery format for the Institute of

Physics Publishing's serial *Modelling and Simulation in Materials Science and Engineering*, but it was eventually delivered in a SGML-based format, which was then converted to HTML for viewing via a Web browser.⁴⁸ The Chemistry Online Retrieval Experiment (CORE), which gave access to American Chemical Society (ACS) serials, converted data from the native format used by the ACS to a variant of the SGML Document Type Definition (DTD) produced by the American Association of Publishers for their Electronic Manuscript Standard.⁴⁹ Articles sometimes exist in a format that can be converted into other formats for delivery to the end-user. For example, a study undertaken for the NEDLIB (Networked European Deposit Library) project found that two thirds of the publishers in a non-randomly selected sample generated HTML "on the fly" from SGML or XML encoded text.⁵⁰ There is still a problem that logical content based mark-up (unlike PostScript or PDF) is still not very good at representing mathematical or chemical information. In order to address these requirements, the World Wide Web Consortium (W3C) is developing a Mathematics Markup Language (MathML) and chemists have published an XML-based Chemical Markup Language (CML) for molecular information.⁵¹

While PDF and HTML versions of articles often replicate the functionality and sometimes appearance of papers in printed serials, the form of articles that can be published in e-serials is not fixed. An electronic serial can be a dynamic document including embedded multimedia or active links to related publications or data. They can also be regularly updated to take account of comments made by scholars in reviews or in other publications. Murray-Rust has argued that in chemistry the "static,

immutable, non-interactive paper publication" is often unable to communicate the real message of an author and that a publication could also form a data resource for analysis or for input into programs or instruments.⁵² Experiments with more "dynamic" types of e-serial include *Internet Archaeology* and the CLIC electronic journal project; both initially funded by the UK Electronic Libraries (eLib) programme.⁵³ For example, the CLIC project attempted to develop ways in which users could acquire three-dimensional molecular data in digital form through electronic serials that could act as a starting point for their own exploration of the content.⁵⁴ These new types of publication model will present a severe challenge for preservation. As some of the CLIC researchers have asked, how long should any given data be expected to reside in automatically accessible form on the Internet?⁵⁵ Also, can this data be preserved in such a way that it can be retrieved in the future without also using the specialised software tools that have been developed to process them? These are the types of issues that will have to be considered by those who have responsibility for the long-term preservation of e-serials.

E-serials and digital preservation

The root of the digital preservation problem is technological, but any proposed solution will also need to take account of organisational and economic issues. The following sections will introduce the main technological problems with the long-term preservation of information in digital form, delineate some possible preservation strategies and then outline some of the non-technological issues.

Technological preservation issues

Almost all kinds of digital information need to be interpreted by machines before it can become intelligible to humans. There are preservation problems associated with three distinct aspects of digital information technologies: the medium upon which digital information is recorded and its associated hardware and software.

Media longevity - digital information is mostly stored on either magnetic or optical media types. Both have relatively short lifetimes in comparison with more established media like paper and microfilm. Hedstom argues that the preservation threat posed by magnetic and optical media is "qualitatively different" in that the media are easily reusable and in the fact they deteriorate in a matter of years, not decades.⁵⁶ One immediate technological response to this problem is known as "refreshing" - the periodic recopying of the data bits onto a new medium. The focus of digital preservation has in recent years moved away from media longevity issues, not because the problems have been solved to any great extent, but because there is a greater awareness of the significant technological problems associated with hardware and software obsolescence.

Hardware obsolescence and software dependence - Mallinson noted back in 1987 that one of the most serious problems with preserving electronic information was the rapid obsolescence of electronic hardware.⁵⁷ Brichford and Maher sum up this problem when they say that a "twenty-year life for the plastic backing material used for computer tapes and disks is irrelevant if the tape or disk drives on which they were recorded become obsolete and unavailable after ten years."⁵⁸ In addition, digital

information is often stored in formats that are dependent upon particular software to interpret them correctly. This is known as software dependence.

Digital preservation strategies

It is now becoming clear that the successful long-term preservation of digital information will be dependent upon relevant organisations identifying and implementing suitable preservation strategies.⁵⁹ At present, there are three main technical options: technology preservation, software emulation and data migration. None of these options provide a perfect solution for all types of digital resource and, as Ross notes, selecting any one strategy will require trade-offs to be made.⁶⁰

Technology preservation is the preservation of an information object together with all of the hardware and software needed to interpret it. This approach has its supporters and may have an important short-term role for the recovery of data from obsolete storage media and platforms, but it is unlikely become a viable long-term strategy.⁶¹ Feeney points out that collection managers who relied only upon this approach would soon end up with "a museum of ageing and incompatible computer hardware."⁶² As a result, the other proposed strategies for digital preservation are not concerned with the preservation of physical artefacts (hardware, media, etc.), but concentrate instead upon the preservation of the information objects themselves in some disembodied digital form.⁶³ Both emulation and migration strategies are examples of this general approach.

Emulation strategies are based on the premise that the best way to preserve the functionality and "look-and-feel" of digital resources would be to preserve the original software and then to run this on emulators that would mimic the behaviour of the obsolete hardware and operating systems.⁶⁴ Emulation strategies would require the encapsulation of a digital data object together with the application software used to create or interpret it and a description of the required hardware environment that would be a specification for an emulator. While preservation strategies based on emulation may be promising for helping to preserve the look and feel of complex objects - e.g. interactive molecular diagrams - they are also likely to be very complicated and, therefore, expensive to implement.⁶⁵

Migration strategies are more tried-and-tested, and mean the periodic transfer of digital information from one generation of computer technology to a subsequent one.⁶⁶ The point of migration is to transfer to new formats while, wherever possible, preserving the integrity of the information. The simplest migration strategies would involve transfer into a standardised format that would be relatively software independent, e.g. plain ASCII text. Care would have to be taken with this approach. For example, ASCII might be an adequate format for e-serials consisting only of textual information, but its use would result in a considerable loss of functionality for most STM-based electronic serials or those publications that have tried to incorporate more dynamic features. A related strategy would be to migrate resources to a small number of "standard" formats on their ingestion into a digital repository. For e-serials, these formats could include, for example, HTML, PDF or documents structured in

SGML or XML together with their associated DTDs. This strategy would simplify future migration processes while helping to maintain some of the important characteristics of the original.⁶⁷ However, as Ross points out, data migration inevitably leads to some losses in functionality, accuracy, integrity and usability.⁶⁸ Successful migration strategies will, therefore, depend upon metadata being created to record the migration history of a digital object and to record contextual information so that future users can either reconstruct or - at the very least - begin to understand the technological environment in which a particular digital object was created. In general, e-serials would be good candidates for migration strategies, except where they contain significant amounts of multimedia.

One publisher of e-serials has already announced a migration-based plan for the long-term preservation of its own Web-based titles. In February 2000, HighWire Press described how content supplied from publishers would be migrated to industry-standard formats and then stored on robust disk arrays in multiple locations.⁶⁹ These would then be migrated as necessary. Together with Sun Microsystems, Stanford University Libraries are also involved in the LOCKSS (Lots of Copies Keeps Stuff Safe) program.⁷⁰ This involves the development of open-source software that would enable the content of e-serials to be cached on a large number of distributed sites. The LOCKSS system then automatically and continuously monitors these distributed caches and repairs files when it finds that their content has been damaged or removed.

Preservation metadata

Recent developments have suggested that, regardless of which particular digital preservation strategy is chosen, the successful long-term preservation of digital resources will depend upon the creation and maintenance of metadata that will be able to record some part of a digital object's functionality and context.⁷¹ Lynch says that, within a digital repository, "metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information."⁷²

Much of the current focus on preservation metadata has been centred around the *Reference Model for an Open Archival Information System (OAIS)* being developed by the Consultative Committee for Space Data Systems (CCSDS) on behalf of the International Organization for Standardization (ISO).⁷³ The OAIS model defines a range of functions that are applicable to any archive, whether digital or not. These functions include ingest, archival storage, data management, administration, and access. Amongst other things, the OAIS model aims to provide a common framework that can be used to help understand archival challenges and especially those that relate to digital information. Accordingly, the OAIS model identifies and distinguishes between the various types of metadata that will need to be recorded by a digital repository. Several digital library projects have begun to use the OAIS model as a means of helping to define digital preservation systems and their associated metadata. These include NEDLIB, which has developed a model for a

deposit system for electronic publications (DSEP),⁷⁴ and the UK Cedars (CURL Exemplars in Digital Archives) project.⁷⁵

Non-technological preservation issues

Despite their apparent intractability, technological problems are probably not the most important factor in the preservation of e-serials. Strategies can be implemented if the need for preservation is noticed early enough. The following sections will introduce some of the most important non-technological issues that will also need to be addressed to ensure the long-term preservation of scholarly information published in e-serials.

Authenticity and intellectual preservation

Perhaps the most important preservation requirement for scholarly communication in digital form is for its integrity to be guaranteed. In comparison with printed information, online digital information is relatively easy to change and update. Indeed, this is one of its major advantages over printed information. However, with preservation in mind, this characteristic becomes a potential problem. It is very difficult to prove that digital information has not been accidentally or deliberately corrupted at some time. It is also very difficult to know exactly when a particular version of a paper was published. This has implications for the integrity of the scholarly record and for the retrospective establishment of priority over particular discoveries or advances. In the digital world there is scope to frequently update scholarly and scientific papers to take account of new data, more recent research and

the comments of other scholars. For some, this is one of the major advantages of digital publication but, as Lynch notes, this is "culturally opposed to the view of the scholarly record as comprising a series of discrete, permanently fixed contributions of readily attributable authorship."⁷⁶

In short, the users of e-serials will want to be sure that the version of the paper that they are referring to is the one that they want to see and that it has not been deliberately or accidentally corrupted since its original publication date. With relation to electronic serials, authenticity could be promoted by the adoption of techniques based on cryptographic theory. Graham has suggested utilising technologies like digital time-stamping (DTS), which uses one-way cryptographic hashing techniques,⁷⁷ but recognises that there are likely to be other solutions.⁷⁸

The importance of intellectual preservation for the future of scholarly communication cannot be over-emphasised. Without some enduring way of ensuring the authenticity of digital documents over time, it is possible that networks will ultimately be unable to support some of the basic requirements of scholarly communication.

Licensing and intellectual property rights

The growth in provision and use of digital information resources (including e-serials published by commercial or learned society publishers) has resulted in fundamental changes in the way information is owned. The information content of printed serials has been usually purchased outright by subscription, either directly

from the publishers or through subscription agents. The organisation or individual that purchases a serial will then normally retain physical custody of the artefact (e.g., a volume or issue) for as long as it is required. Assuming that this artefact is kept in an appropriate environment and safeguarded against disaster (e.g., fire and flood), it should last for a long time. Given the fact that printed serials will normally be subscribed to by more than one organisation, a distributed "canonical archive" of scholarly serials will be built up in this way. Long-term preservation and access is essentially a by-product of this process and does not require specific initiation.

The situation of digital information resources is quite different. The "purchaser" of a digital resource like an e-serial (unless it is a physical artefact like a CD-ROM) does not necessarily retain "physical custody" over it. Concerns over copyright mean that commercial publishers tend to license the use of information to customers, thus ensuring contract law rather than copyright law governs the use of the information.⁷⁹ A license for a commercial electronic serial would normally only give a "subscriber" specific rights over the use of a particular serial, or group of serials, for a limited amount of time. This is where the position becomes problematic. What happens when the subscribing institution decides to cancel its subscription? Would all access rights to that serial, including those back-issues already "paid for," then be removed? What would happen if the serial itself is no longer published or if the publisher is taken over or ceases trading? Research organisations and libraries might then find that they have no direct control over which particular e-serials can be preserved as part of the "canonical archive" of scholarly communication. Because

most licenses are time-limited, Okerson has argued that they are a flexible way of ensuring that libraries' concerns about ongoing access can be addressed in the agreements negotiated between libraries and publishers.⁸⁰ However, if publishers commit to providing continued access to serial back issues after termination of the license, this may have cost implications for the publisher.⁸¹ It is possible that some digital publications will disappear before all licensing problems are resolved.

Responsibility for preservation

Borgman has written that scholars, as both authors and readers, "expect access to publications long after the documents' authors are dead, the serials in which they appeared have ceased to be published, or the publishers have been acquired by an international conglomerate or otherwise ceased to exist."⁸² In the print world, the task of preserving the published scholarly record has historically fallen to research libraries. In a world of e-serials and e-print archive services it is by no means clear that libraries will have exactly the same role. Who, then, should be responsible for preserving the scholarly record in digital form?

Several different answers are possible. Some, including the CPA/RLG task force, argued for a decentralised approach. The task force report suggested that digital information creators, providers and owners should have the initial responsibility for archiving while certified digital archives would be given the legal rights and duties to exercise an aggressive fail-safe rescue function.⁸³ For serials, this suggests that the

responsibility for the preservation of the scholarly record would lie initially with the publishers of e-serials or with the providers of institutional e-print archives.

In the short term, some commercial and learned society publishers might welcome a decentralised approach because they would have a commercial interest in controlling user-access to the content of e-serials. In this scenario, publishers would maintain and migrate the digital versions of papers themselves until such time as the serial is no longer of economic value.⁸⁴ So, for example, the storage hardware might need to be periodically upgraded, sometimes delivery formats would change, but the serial itself need never go "out of print." The most critical time will come when the publishers no longer have any economic interest in keeping a serial going, or when the publisher is taken over or otherwise ceases trading. It is at this point that some kind of fail-safe mechanism would need to be activated. This type of approach, inevitably, would include a level of risk and would require good leadership and co-ordination from information professionals.⁸⁵ There is some evidence to suggest that some publishers are aware of their preservation responsibilities and are sensitive to libraries concerns.⁸⁶ Elsevier Science, for example, declare in their licenses their intention to maintain the digital files of ScienceDirect in perpetuity, and commit themselves to transferring them to another depository if they are unable to do so.⁸⁷

A second and more traditional approach would be to adapt the more centralised models that have been used to maintain the existing printed scholarly record. Under this model, national libraries and some of the larger research libraries store and maintain large collections of printed serials. In order to perpetuate this

system, many national libraries use some kind of legal deposit legislation to ensure that all relevant published works are collected and preserved. A few countries have successfully extended legal deposit legislation to cover digital publications, but this does not always include online publications like e-serials. This situation, however, is slowly changing. For example, the report of a recent UK Working Party on Legal Deposit recommended that e-serials that have additional information content beyond the print original and those that are available only in digital form should be considered as suitable candidates for future legal deposit legislation.⁸⁸

The products of self-archiving initiatives offer some different challenges. In the same way as serial publishers, centralised services like the Los Alamos e-print archive or PubMed Central definitely do have an interest in maintaining access to the papers that are stored on their systems.⁸⁹ Some advocates of self-archiving go further. One of the Public Library of Science group's justifications for insisting that publishers transfer digital copies of papers to services like PubMed Central after six months is that such services will be able continue the preservation role traditionally undertaken by research libraries. Little thought appears to have been made about the long-term implications of this. The editor of *Academic Medicine* has argued that many of those who have called for services like PubMed Central have, "not recognised the expertise and long-term difficulty of assuming the role previously played by libraries as the science community's archivists."⁹⁰ The position is even more problematic where authors self-archive papers only on institutional e-print servers. The institutions that host such services will somehow need to be made aware that they have the primary

responsibility for the long-term preservation of all papers stored in the service. It is far from clear at the moment how this could be co-ordinated.

In practice, the system that is most likely to evolve will neither be completely centralised nor decentralised. Preservation in the short term will probably become part of the responsibility of publishers or of some centralised e-print services, as well as becoming part of the continuing role of some national and research libraries. New types of third party services may also develop, and Anderson has suggested that librarians should attempt to shape their development.⁹¹ One danger is that some players in the scholarly communication process may just assume that existing libraries will be able to continue to preserve scholarly communication without first ensuring that they have the necessary technical, legal and financial means to do so. Another possible problem is that many preservation decisions might be taken with regard to short-term commercial interests rather than in the wider public interest.⁹² There is, therefore, a need to make all participants in the scholarly communication process more aware of digital preservation issues. In addition, there will be a need for effective co-operation between publishers, libraries and all of the other stakeholders in the scholarly communication process. Boyce says that all of these will have to work together in order to preserve access to digital information.⁹³ Organisations like the Digital Library Federation (DLF) and the Council for Library and Information Resources (CLIR) in the US and the UK Digital Preservation Coalition may have a role in helping to foster this co-operation.⁹⁴

One significant recent development has been the development of the Andrew W. Mellon Foundation's e-Journal archiving program. The Foundation is funding several major US libraries to undertake the development of practical e-journal repositories that meet specific requirements identified by the DLF together with CLIR and the Coalition for Networked Information (CNI). While some of the participating libraries are concentrating on the development of tools or solutions for dynamic serials, Yale, Harvard and the University of Pennsylvania have chosen to work with individual publishers. For example, Yale University Library has announced that it will collaborate with Elsevier to create a digital repository for all of the serials published by Elsevier Science.⁹⁵ These types of initiatives provide models for publisher and library co-operation, as well as offering the opportunity for some practical experiences with digital preservation.

Selection for preservation

Deciding what needs to be preserved could be another problem. The temptation may be - as digital storage devices become increasingly cheaper and more compact - to suggest that everything should be kept. Little is known about the economics of digital preservation over a long period of time but both emulation and migration strategies are likely to be expensive both in terms of time and technical expertise. In normal circumstances, some kind of selection before preservation will be necessary.

For scholarly electronic serials, this process may be made easier by the reliability mechanisms that already exist, e.g. peer-review. Peer-reviewed serials would be obvious candidates for preservation, as would less formal, newsletter-type, publications issued by learned societies and other organisations. However, it may be harder to assess articles "published" in e-print services. The nature of digital information also means that any such appraisal would have to take place very early in its life cycle. It will not be useful to wait until the information has become unavailable before deciding on preservation, or no copies may remain in existence to be preserved. Instead, the identification of suitable resources for preservation should have to be made close to its issue (i.e. publication) date, and rights for its long-term preservation would also need to be negotiated at the same time. In addition, it may be useful in some circumstances if selection and appraisal could become continuous processes.

Another problem is that there is no clear idea as to which particular file format should be seen as the definitive version for preservation. Many e-serial publishers have created SGML-based files that are used to generate PDF or HTML versions for delivery over the Web. Which particular one of these formats should be preserved? Also, where printed versions also exist, should these be kept as well? What about those papers that include embedded multimedia, software and data sets, or contain hyperlinks to other documents, or are otherwise dynamic? What preservation strategies should be adopted for these types of information? Decisions on these issues will have to be made on a title-to-title or, possibly, a paper-to-paper basis.

Conclusion

This paper has attempted to outline some of the problems that will need to be confronted to ensure the continued existence and accessibility of the information content published in scholarly e-serials. There are other important issues that have not been discussed here, most notably the likely economic implications of long-term digital preservation. Librarians, publishers, archivists, computer scientists and others with an interest in digital preservation are currently investigating many of these issues. It is important to realise that many of the problems will only be solved with practical experience of ensuring the preservation of digital information and with practical co-operation between publishers, libraries, scholars and other stakeholders in scholarly communication.

One more point needs to be considered. In the past, the preservation of the canonical archive of scholarly communication was an activity that was considered a specialist activity within the library and information professions. With the advent of digitally based scholarly communication, digital preservation might become one of the basic functions of the digital research library. Indeed, all other digital library activities, including resource discovery and access, may ultimately become dependent upon interested parties solving the digital preservation conundrum.

References

¹ F. W. Lancaster, *Toward Paperless Information Systems* (New York: Academic Press, 1978).

² Donald Waters and John Garrett, eds., *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (Washington, D.C.: Commission on Preservation and Access, 1996), <http://www.rlg.org/ArchTF/> (viewed May 14, 2001).

³ Michael Alexander, "Virtual Stacks: Storing and Using Electronic Journals," *Serials* 10, no. 2 (1997): 173-178.

⁴ Peter Leggate and Mike Hannant, "The Archiving of Online Journals," *Learned Publishing* 13, no. 4 (2000): 246-250.

⁵ John Feather, *Preservation and the Management of Library Collections* (London: Library Association Publishing, 1991), 2.

⁶ Margaret Hedstrom, "Digital Preservation: a Time Bomb for Digital Libraries," *Computers and the Humanities* 31 (1998): 189-202.

⁷ Charles M. Dollar, *Archival Theory and Information Technologies: the Impact of Information Technologies on Archival Principles and Methods* (Ancona, Italy: University of Macerata, 1992), 66.

⁸ David Bearman, *Archival Methods*, Archives and Museum Informatics Technical Report, vol. 3, no. 1 (Pittsburgh, Pa.: Archives and Museum Informatics, 1989), 17-27.

- ⁹ Maurice B. Line, "The Publication and Availability of Scientific and Technical Papers: an Analysis of Requirements and the Suitability of Different Means of Meeting Them," *Journal of Documentation* 48, no. 2 (1992): 201-219.
- ¹⁰ Andrew M. Odlyzko, "Tragic Loss or Good Riddance? The Impending Demise of Traditional Scholarly Journals," *International Journal of Human-Computer Studies* 42 (1995): 71-122.
- ¹¹ Fytton Rowland, "Who Will Buy My Bells and Whistles? The True Needs of Users of Electronic Journals," *Serials* 13, no. 2 (2000): 73-77.
- ¹² Dennis P. Carrigan, "Research Libraries Evolving Response to the 'Serials Crisis'," *Scholarly Publishing* 23, no. 3 (1992): 138-151.
- ¹³ Michael Buckingham, "Where is the Revolution?" *Nature* 311 (27 Sep. 1984): 309-310.
- ¹⁴ Brian Shackel, *BLEND-9: Overview and Appraisal*, British Library Research Paper, 82 (London: British Library, 1991).
- ¹⁵ Fytton Rowland, Cliff McKnight, Jack Meadows and Peter Such, "ELVYN: the Delivery of an Electronic Version of a Journal from the Publisher to Libraries," *Journal of the American Society for Information Science* 47, no. 9 (1996): 690-700.
- ¹⁶ Cliff McKnight, "Electronic Journals: What do Users Think of Them?" Paper presented at ISDL '97: International Symposium on Research, Development

and Practice in Digital Libraries 1997, Tsukuba, Ibaraki, Japan (Nov. 18-21, 1997), <http://www.dl.ulis.ac.jp/ISDL97/proceedings/mcknight.html> (viewed May 14, 2001).

¹⁷ Fytton Rowland, "Print Journals: Fit for the Future?" *Ariadne* 7 (Jan. 1997): 6-7, <http://www.ariadne.ac.uk/issue7/fytton/> (viewed May 14, 2001).

¹⁸ Fytton Rowland, Cliff McKnight and Jack Meadows, eds., *Project ELVYN: an Experiment in Electronic Journal Delivery* (London: Bowker-Saur, 1995).

¹⁹ Clifford Lynch, "The TULIP project: context, history and perspective," *Library Hi Tech* 13, no. 4 (1995): 8-24.

²⁰ Charles Oppenheim, Clare Greenhalgh and Fytton Rowland, "The Future of Scholarly Journal Publishing," *Journal of Documentation* 56, no. 4 (2000): 361-398.

²¹ Kenneth Frazier, "The Librarians' Dilemma: Contemplating the Costs of the 'Big Deal'," *D-Lib Magazine* 7, no. 3 (Mar. 2001), <http://www.dlib.org/dlib/march01/frazier/03frazier.html> (viewed May 14, 2001).

²² Leah Halliday and Charles Oppenheim, "Developments in Digital Journals," *Journal of Documentation* 57, no. 2 (2001): 260-283.

²³ Robert Bley, "NESLI: a Successful National Consortium," *Library Consortium Management: an International Journal* 2, no. 1 (2000): 18-28.

²⁴ Sarah C. Michalak, "The Evolution of SPARC," *Serials Review* 26, no. 1 (2000): 10-21.

²⁵ Ronald E. LaPorte, Eric Marler, Shunichi Akazawa, Francois Sauer, Carlos Gamboa, Chris Shenton, Caryle Glosser, Anthony Villasenor and Malcolm Maclure, "The Death of Biomedical Journals," *BMJ* 310 (May 27, 1995): 1387-1390.

²⁶ Stevan Harnad and Matt Hemus, "All or None: no Stable Hybrid or Half-Way Solutions for Launching the Learned Periodical Literature into the Post-Gutenberg Galaxy," in *The Impact of Electronic Publishing on the Academic Community* ed. Ian Butterworth (London: Portland Press, 1998), <http://tiepac.portlandpress.co.uk/books/online/tiepac/session1/ch5.htm> (viewed May 14, 2001).

²⁷ Ann Okerson and James O'Donnell eds., *Scholarly Journals at the Crossroads: a Subversive Proposal for Electronic Publishing* (Washington, D.C.: Association of Research Libraries, 1995), <http://www.arl.org/scomm/subversive/index.html> (viewed May 14, 2001).

²⁸ Harnad and Hemus, "All or None."

²⁹ Stevan Harnad and Jessie Hey, "Esoteric Knowledge: the Scholar and Scholarly Publishing on the Net," in *Networking and the Future of Libraries 2: Managing the Intellectual Record*, ed. Lorcan Dempsey, Derek Law and Ian Mowat (London: Library Association Publishing, 1995), 110-116.

³⁰ Stevan Harnad, "The Self-Archiving Initiative," *Nature* 410 (26 Apr. 2001): 1024-1025.

³¹ Paul Ginsparg, "First Steps Towards Electronic Research Communication," *Computers in Physics* 8, no. 4 (1994): 390-396.

³² Steven B. Giddings, quoted in Gary Stix, "The Speed of Write," *Scientific American* 271, no. 6 (Dec. 1994): 72-77.

³³ Harold Varmus, *PubMed Central: an NIH-Operated Site for Electronic Distribution of Life Sciences Research Reports* (Bethesda, Md.: National Institutes of Health, Aug. 30, 1999), <http://www.nih.gov/welcome/director/pubmedcentral/pubmedcentral.htm> (viewed May 14, 2001).

³⁴ Arnold S. Relman, "The NIH 'E-biomed' Proposal - a Potential Threat to the Evaluation and Orderly Dissemination of New Clinical Studies," *New England Journal of Medicine* 340, no. 23 (Jun. 10, 1999): 1828-1829.

³⁵ Edward J. Valauskas, "Waiting for Thomas Kuhn: First Monday and the Evolution of Electronic Journals," *First Monday* 2, no. 12 (Dec. 1997), http://www.firstmonday.dk/issues/issue2_12/valauskas/ (viewed May 14, 2001).

³⁶ Andrew Odlyzko, "The Economics of Electronic Journals," *First Monday* 2, no. 8 (Aug. 1997), http://www.firstmonday.dk/issues/issue2_8/odlyzko/ (viewed May 14, 2001).

³⁷ Richard J. Roberts, Harold E. Varmus, Michael Ashburner, Patrick O. Brown, Michael B. Eisen, Chaitan Khosla, Marc Kirchner, Roel Nusse, Matthew

Scott and Barbara Wold, "Building a 'GenBank' of the Published Literature," *Science* 291 (Mar. 23, 2001): 2318-2319.

³⁸ "Public Library of Science," <http://www.publiclibraryofscience.org/> (viewed May 14, 2001).

³⁹ Stevan Harnad, "On-line Journals and Financial Fire Walls," *Nature* 395 (Sep. 10, 1998): 127-128.

⁴⁰ Christine L. Borgman, *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World* (Cambridge, Mass.: MIT Press, 2000), 91.

⁴¹ Jack Meadows, "Scholarly Communication and the Serial," in *Scholarly Communication and Serials Prices*, ed. Karen Brookfield (London: Bowker-Saur, 1991), 5-14.

⁴² Barrie T. Stern and Henk C. J. Compier, "ADONIS: Document Delivery in the CD-ROM Age," *Interlending and Document Supply* 18, no. 3 (1990): 79-87.

⁴³ Steve Hitchcock, Leslie Carr and Wendy Hall, "A Survey of STM Online Journals, 1990-95: the Calm Before the Storm," in *ARL Directory of Electronic Journals, Newsletters and Academic Discussion Lists*, 6th ed. (Washington, D.C.: Association of Research Libraries, 1996), 7-32, <http://journals.ecs.soton.ac.uk/survey/survey.html> (viewed May 14, 2001).

⁴⁴ Judith Wusteman, "Electronic Journal Formats," *Program* 30, no. 4 (1996): 319-343.

⁴⁵ Mark Bide & Associates, *Standards for Electronic Publishing: an Overview*, NEDLIB Report Series, 3 (The Hague: Koninklijke Bibliotheek, 2000), 7.

⁴⁶ Paul Tyers, "Roman Amphoras in Britain," *Internet Archaeology* 1 (1996), http://intarch.ac.uk/journal/issue1/tyers_index.html (viewed May 14, 2001).

⁴⁷ Richard Entlich, Lorrin Garson, Michael Lesk, Lorraine Normore, Jan Olsen and Stuart Weibel, "Testing a Digital Library: User Response to the CORE Project," *Library Hi Tech* 14, no. 4 (1996): 99-118.

⁴⁸ Cliff McKnight, Jack Meadows, David Pullinger and Fytton Rowland, *ELVYN: Publisher and Library Working towards the Electronic Distribution and Use of Journals*. Paper presented at Digital Libraries '94: the First Annual Conference on the Theory and Practice of Digital Libraries, College Station, Texas (Jun. 19-21, 1994), <http://csdl.tamu.edu/DL94/paper/mcknight.html> (viewed May 14, 2001).

⁴⁹ Michael E. Lesk, "Electronic Chemical Journals," *Analytical Chemistry* 66, no. 14 (1994): 747A-755A.

⁵⁰ Mark Bide & Associates, *Standards for Electronic Publishing*, 17.

⁵¹ Peter Murray-Rust and Henry S. Rzepa, "Chemical Markup, XML, and the Worldwide Web: 1. Basic Principles," *Journal of Chemical Information and Computer Sciences* 39, no. 6 (1999): 928-942.

⁵² Peter Murray-Rust, "The Globalization of Crystallographic Knowledge," *Acta Crystallographica D* 54 (1998): 1065-1070.

⁵³ CLIC was an acronym made up from the four organisations involved in the project: the Royal Society of Chemistry, the University of Leeds, Imperial College of Science, Technology and Medicine, London and the University of Cambridge.

⁵⁴ David James, Benjamin J. Whitaker, Christopher Hildyard, Henry S. Rzepa, Omer Casher, Jonathan M. Goodman, David Riddick and Peter Murray-Rust, "The Case for Content Integrity in Electronic Chemistry Journals: the CLIC Project," *New Review of Information Networking* 1 (1995): 61-69.

⁵⁵ Omer Casher, Gudge K. Chandramohan, Martin J. Hargreaves, Christopher Leach, Peter Murray-Rust, Henry S. Rzepa, Roger Sayle and Benjamin J. Whitaker, "Hyperactive Molecules and the World-Wide-Web Information System," *Journal of the Chemical Society: Perkin Transactions 2*, no 1 (Jan. 1995): 7-11.

⁵⁶ Margaret Hedstrom, "Preserving the Intellectual Record: a View from the Archives," in *Networking and the Future of Libraries 2: Managing the Intellectual Record*, ed. Lorcan Dempsey, Derek Law and Ian Mowat (London: Library Association Publishing, 1995), 180.

⁵⁷ John C. Mallinson, "On the Preservation of Human- and Machine Readable Records," *Information Technology and Libraries* 7 (1988): 19-23.

⁵⁸ Maynard Brichford and William Maher, "Archival Issues in Network Electronic Publications," *Library Trends* 43, no. 4 (1995): 701-712.

⁵⁹ Neil Beagrie and Daniel Greenstein, *A Strategic Policy Framework for Creating and Preserving Digital Collections* (London: South Bank University, Library Information Technology Centre, 1998), <http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html> (viewed May 14, 2001).

⁶⁰ Seamus Ross, "Consensus, Communication and Collaboration: Fostering Multidisciplinary Co-operation in Electronic Records," in *Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996* (Luxembourg: Office for Official Publications of the European Communities, 1997), 330-336.

⁶¹ Seamus Ross and Ann Gow, *Digital Archaeology: Rescuing Neglected and Damaged Data Resources* (London: South Bank University, Library Information Technology Centre, 1999).

⁶² Mary Feeney, ed., *Digital Culture: Maximising the Nation's Investment* (London: National Preservation Office, 1999), 42.

⁶³ Cliff Lynch, "Canonicalization a Fundamental Tool to Facilitate Preservation and Management of Digital Information," *D-Lib Magazine* 5 no. 9, (Sep. 1999), <http://www.dlib.org/dlib/september99/09lynch.html> (viewed May 14, 2001).

⁶⁴ Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* (Washington, D.C.: Council on Library and Information Resources, 1999), <http://www.clir.org/pubs/abstract/pub77.html> (viewed May 14, 2001).

⁶⁵ David Bearman, "Reality and Chimeras in the Preservation of Electronic Records." *D-Lib Magazine* 5, no. 4 (Apr. 1999), <http://www.dlib.org/dlib/april99/bearman/04bearman.html> (viewed May 14, 2001).

⁶⁶ Waters and Garrett, *Preserving Digital Information*, 6.

⁶⁷ Margaret Hedstrom, "Preserving the Intellectual Record," 185.

⁶⁸ Ross, "Consensus, Communication and Collaboration," 331.

⁶⁹ Stanford University Libraries, *HighWire Press Ensures that Online Publications don't get Lost in Cyberspace* (Stanford, Calif.: Stanford University, Feb. 4, 2000), <http://www.stanford.edu/dept/news/pr/00/000204highwire.html> (viewed May 14, 2001).

⁷⁰ Vicky Reich and David S. H. Rosenthal, "LOCKSS (Lots of Copies Keep Stuff Safe)," *New Review of Academic Librarianship* 6 (2000): 155-161.

⁷¹ Michael Day, "Issues and Approaches to Preservation Metadata," in *Guidelines for Digital Imaging* (London: National Preservation Office, 1999), 73-84, <http://www.rlg.org/preserv/joint/day.html> (viewed May 14, 2001).

⁷² Cliff Lynch, "Canonicalization."

⁷³ Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-R-1 (1999), http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html (viewed May 14, 2001).

⁷⁴ Titia van der Werf, *The Deposit System for Electronic Publications: a Process Model*, NEDLIB Report Series, 6 (The Hague: Koninklijke Bibliotheek, 2000).

⁷⁵ Kelly Russell, Derek Sergeant, Andy Stone, Ellis Weinberger and Michael Day, *Metadata for Digital Preservation: the Cedars Project Outline Specification* (Leeds: Cedars Project, 2000), <http://www.leeds.ac.uk/cedars/metadata.html> (viewed May 14, 2001).

⁷⁶ Clifford A. Lynch, "Integrity Issues in Electronic Publishing," in *Scholarly Publishing: the Electronic Frontier*, ed. Robin P. Peek and Gregory B. Newby (Cambridge, Mass.: MIT Press, 1996), 133-145.

⁷⁷ Hashing techniques use algorithms that convert the arrangement of all characters, symbols, graphics, etc. within a particular document into a unique hash value that can be stored and retrieved as metadata. Any change (however small) to the document will produce a different hash value when it is converted using the same algorithm. The process is described as "one-way" because there is no means of recreating the original document from its hash value.

- ⁷⁸ Peter S. Graham, "Long-Term Intellectual Preservation," in *Digital Imaging Technology for Preservation*, ed. Nancy E. Elkington (Mountain View, Calif.: Research Libraries Group, 1994), 41-57.
- ⁷⁹ Ann Okerson, "What Academic Libraries Need in Electronic Content Licenses," *Serials Review* 22, no. 4 (1996): 65-69.
- ⁸⁰ Ann Okerson, Copyright in the Year 2010: No Longer an Issue for Scholarly Electronic Publishing. *Serials Review* 25, no. 4 (1999): 33-35.
- ⁸¹ Sally Morris, "Archiving Electronic Publications: What are the Problems and Who Should Solve Them?" *Serials Review* 26, no. 3 (2000): 64-68.
- ⁸² Christine L. Borgman, "Digital Libraries and the Continuum of Scholarly Communication," *Journal of Documentation* 56, no. 4 (2000): 412-430.
- ⁸³ Waters and Garrett, *Preserving Digital Information*, 21.
- ⁸⁴ Morris, "Archiving Electronic Publications," 67.
- ⁸⁵ Margaret Hedstrom, "Electronic Archives: Integrity and Access in the Network Environment," in *Networking in the Humanities*, ed. Stephanie Kenna and Seamus Ross (London: Bowker-Saur, 1995), 77-95.
- ⁸⁶ Ellen Finnie Duranceau, "Archiving and Perpetual Access for Web-Based Journals: a Look at the Issues and How Five E-journal Providers are Addressing them," *Serials Review* 24, no. 2 (1998): 110-115.

- ⁸⁷ Karen Hunter, "Digital Archiving," *Serials Review* 26, no. 3 (2000): 62-64.
- ⁸⁸ Geoff Smith, "The Legal Deposit of Non-Print Publications: the 1998 Working Party on Legal Deposit," *Serials* 12, no. 2 (1999): 125-129.
- ⁸⁹ Richard E. Luce, "E-prints Intersect the Digital Library: Inside the Los Alamos arXiv," *Issues in Science and Technology Librarianship* 29 (Winter 2001), <http://www.library.ucsb.edu/istl/01-winter/article3.html> (viewed May 14, 2001).
- ⁹⁰ Addeane S. Caellegh, "PubMed Central and the New Publishing Landscape: Shifts and Tradeoffs," *Academic Medicine* 75, no. 1 (2000): 4-10.
- ⁹¹ Rick Anderson, "Is the Digital Archive a New Beast Entirely?" *Serials Review* 26, no. 3 (2000): 50-52.
- ⁹² David Haynes and David Streatfield, "Who will Preserve Electronic Publications?" *Serials* 10, no. 3 (1997): 345-351.
- ⁹³ Peter B. Boyce, "Who Will Keep the Archives? Wrong Question!" *Serials Review* 26, no. 3 (2000): 52-55.
- ⁹⁴ Neil Beagrie, "The JISC Digital Preservation Focus and the Digital Preservation Coalition," *New Review of Academic Librarianship* 6 (2000): 257-267.
- ⁹⁵ Yale University Library, *Yale Library to Plan Digital Archives with Elsevier Science* (New Haven, Conn.: Yale University Library, Feb. 23, 2001),

http://www.library.yale.edu/Administration/newsreleases/elsevier_release.html

(viewed May 14, 2001).